

2019 年 4 月 15 日周报

本周采用 Python 图卷积神经网络的实现，并结合 LSTM 模型进行了预测。因为之前一直没有写周报，特此对这一个月的工作进行总结。本月主要做了一下几点工作：

- 学习 Python 开始，然后搭建 TensorFlow、keras 深度学习框架，了解了其实现方式。
- 对合肥市交通事故数据进行了预处理，处理方式如下：
 - a) 选取处理区域，在整个城市范围内的数据缩小到一环内数据（因数据量变小，大大缩短训练时间，便于测试，也方便构建图网络结构），选取区域如图所示，经度（117.2433，117.3122）维度（31.8478，31.8834）



- b) 交通事故映射，因事故数据集中只有经纬度没有明确的标明在那条道路上发生，因此对事故和道路进行了映射，判读事故发生在哪条道路上。
- c) 判断道路是否相交，构建路网图结构。道路是由道路分段构成的，计算每条道路是否与其他路段相交，入道路相交则在 neighbor 字段中添加记录。

id	xlng	xlat	yling	ylat	neighbor	way
1	117.277588	31.850651	117.280678	31.854868	_2_124_139	金寨路
2	117.280678	31.854868	117.282124	31.856908	_1_3	金寨路
3	117.282124	31.856908	117.282762	31.85792	_2_4	金寨路
4	117.282762	31.85792	117.28322	31.859047	_3_5	金寨路
5	117.28322	31.859047	117.282735	31.863149	_4_6	金寨路

- d) 跟道路相交情况，重新构建划分路段，构建图网络结构
- e) 计算事故所属的道路分段。计算完成后，在事故记录的数据库 gid 字段中记录道路分段 id。结果，如图。

JJDBH	lng	lat	rid	lm	date	gid
2015020514334102	117.29347028313686	31.8653928128844	1	徽州大道	2015-02-05	28
2015020518541712	117.29234056352108	31.852758464944564	1	徽州大道	2015-02-05	27
2015020608320506	117.29247507083052	31.856781607720315	1	徽州大道	2015-02-06	27
2015020609490011	117.29259934146143	31.858834777902906	1	徽州大道	2015-02-06	28
2015020609495802	117.29385469237144	31.86836801938045	1	徽州大道	2015-02-06	29
2015020610101403	117.2922581533461	31.850293558750955	1	徽州大道	2015-02-06	27
2015020610235822	117.29259663297681	31.85881438349833	1	徽州大道	2015-02-06	28
2015020610442101	117.29235547681539	31.853204524809627	1	徽州大道	2015-02-06	27
2015020618395104	117.29235642245217	31.85323280901082	1	徽州大道	2015-02-06	27

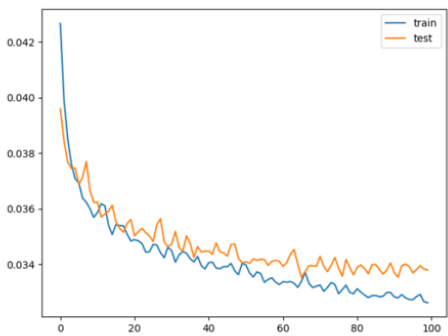
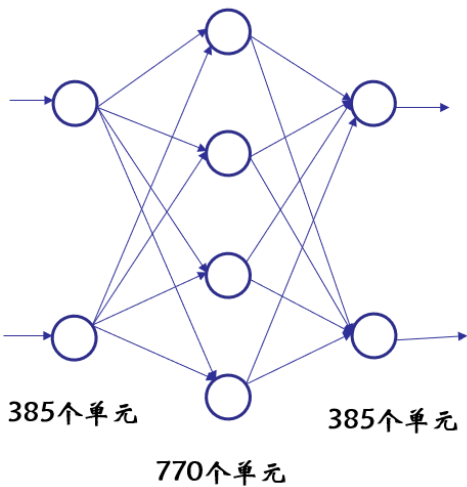
- f) 以天为单位，计算每天发生的交通事故数目。本数据库记录从 2015 年 02 月 04 日至 2018 年 12 月 18 日，共 1414 天记录，本文中，选取前 1000 条为训练数

数据集，后 400 条为测试数据集，数据格式如下图：

JJDBH	lng	lat	rid	lm	date	gid
2015020514334102	117.29347028313686	31.8653928128844	1	徽州大道	2015-02-05	28
2015020518541712	117.29234056352108	31.852758464944564	1	徽州大道	2015-02-05	27
2015020608320506	117.29247507083052	31.856781607720315	1	徽州大道	2015-02-06	27
2015020609490011	117.29259934146143	31.858834777902906	1	徽州大道	2015-02-06	28
2015020609495802	117.29385469237144	31.86836801938045	1	徽州大道	2015-02-06	29
2015020610101403	117.2922581533461	31.850293558750955	1	徽州大道	2015-02-06	27
2015020610235822	117.29259663297681	31.85881438349833	1	徽州大道	2015-02-06	28
2015020610442101	117.29235547681539	31.853204524809627	1	徽州大道	2015-02-06	27
2015020618395104	117.29235642245217	31.85323280901082	1	徽州大道	2015-02-06	27

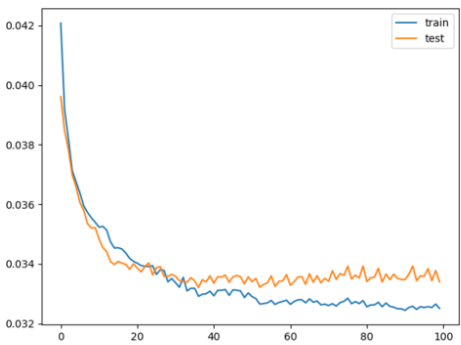
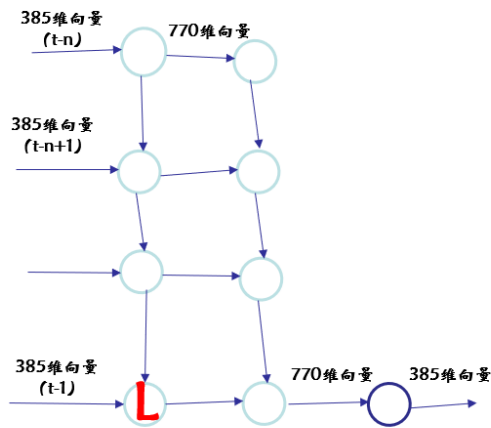
- 采用不同模型进行训练，并展示结果。

a) 简单的全连接网络



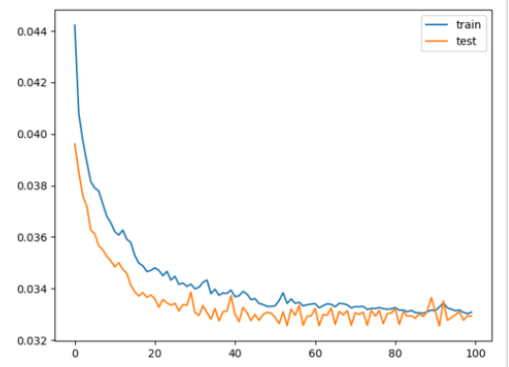
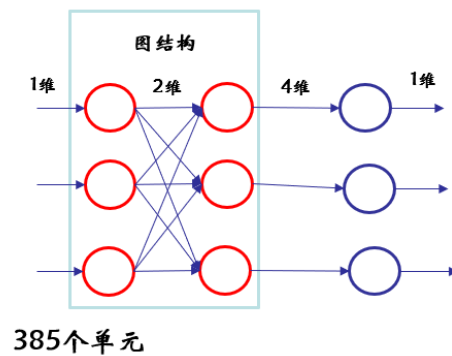
由t-1时刻的事故发生情况，推测t时刻的事故发生情况。

b) 仅采用 LSTM 模型进行预测：



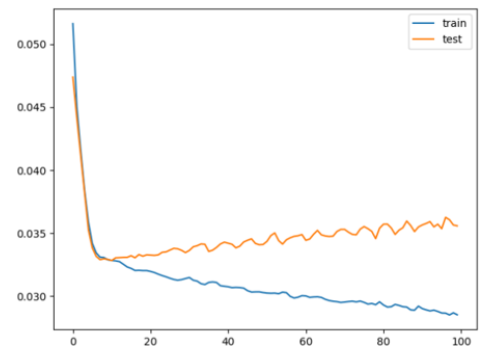
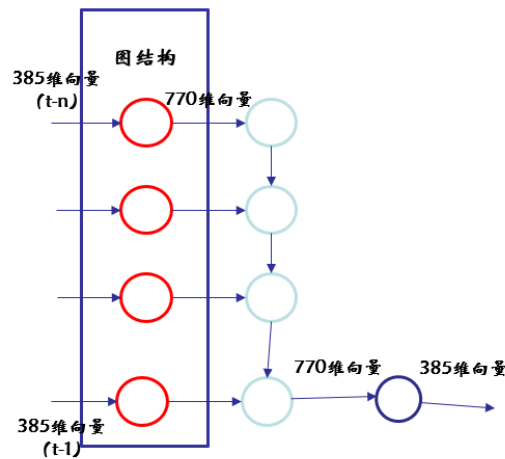
由t-n至t-1时刻的事故发生情况，采用 LSTM模型推测t时刻的事故发生情况。

c) 采用图卷积结构进行预测：



由 $t-1$ 时刻的事故发生情况，采用图卷积模型，推测 t 时刻的事故发生情况。

d) 采用图卷积结构加 LSTM 进行预测



由 $t-n$ 至 $t-1$ 时刻的事故发生情况，采用卷积模型和LSTM模型推测 t 时刻的事故发生情况。

- 对数据进行分析发信采用图卷积结构和 LSTM 结合时可以很好的模拟效果，但存在过拟合现象，导致测试数据误差有上扬倾向。考虑原因有以下：

- 模型参数过多
- 损失函数定义不正确
- 模型的特征维度较少（目前仅采用了事故发生数目作为特征）

现在针对第二点和第三点进行改进，尤其是第三点，考虑要涉及到更多的特征，例如天气，道路限速，道路情况，成流量情况等特征。目前正在整理特征，修改深度学习框架，使其能处理多维度特征。